CD Briefing Meeting
May 14, 2003

LATBauerdick, Ruth Pordes

Open Science Grid

We propose to develop a plan over the next weeks

- and a proposal to the DOE and NSF, over the next few months

and to forge an organization and collaboration

- building upon the previously proposed "Open Science Consortium"

to build an "Open Science Grid" in the US on a "Peta-Scale"

- for the LHC and other science communities

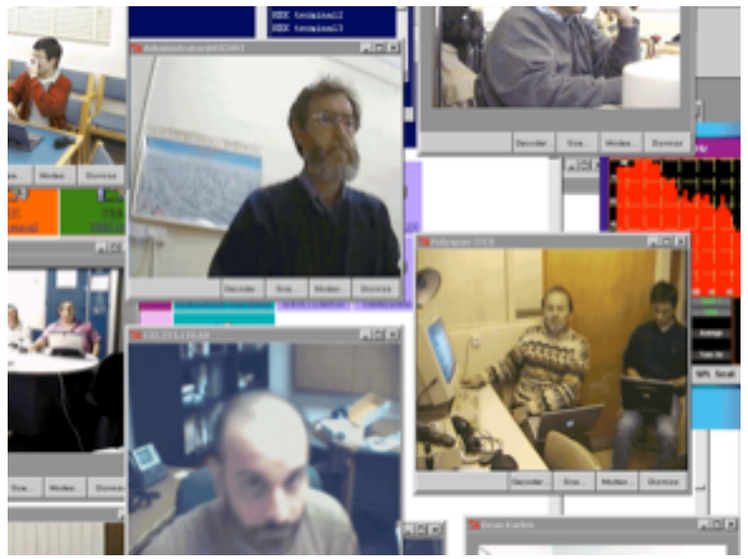
Goals and Scope:

- Develop and deploy services and capabilities for a Grid infrastructure that would make LHC computing resources, and possibly other computing resources for HEP and and other sciences (Run 2 etc) available to the LHC Science community,
  - as a functional, managed, supported and persistent US national resource.
- Provide a persistent 24x7 Grid that peers and interoperates, interfaces to and integrates with, other national and international Grid infrastructures
  - in particular the EGEE in Europe (which will provide much of the LHC Grid resources in Europe to the LCG)

This would change how we do business in US LHS and maybe in Fermilab

LHC Physics Discoveries at Universities and Labs!

U.S. CMS is Committed to Empower the CMS Scientists at U.S. Universities and Labs to do Research on LHC Physics Data



This is why we are pushing Grids and other Enabling Technology

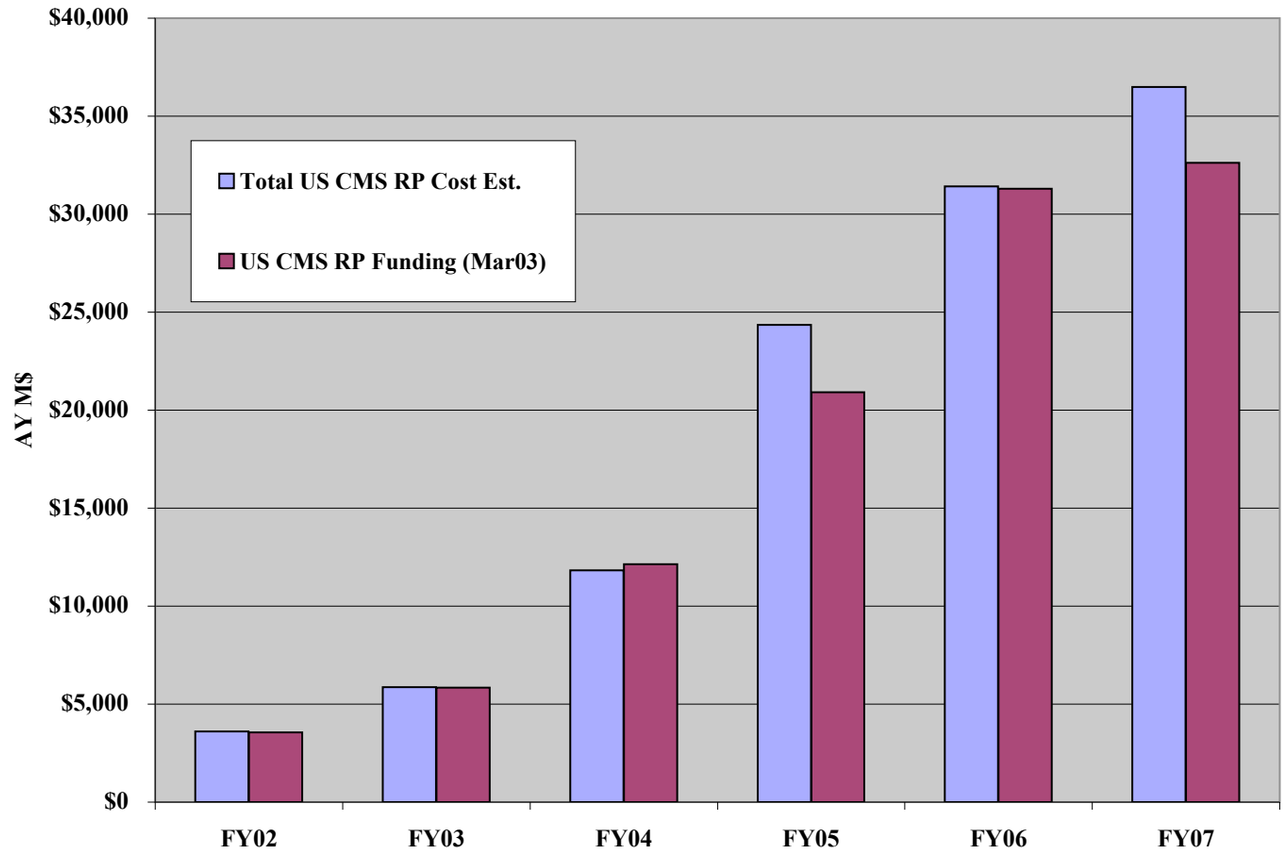
a new, broader, way of thinking about LHC Computing and about what we are constructing with our LHC Computing projects.

- A "Global Village" is a metaphor [...] to describe a place where a community of people live and work and carry out certain activities that are required in order to do science on a global scale.
- We are building an environment for doing science – with a new culture, a new economics of use of computing resources, new levels of both sharing and competitiveness, new partnerships to those outside our field
- Building an environment requires components to be built or found and pieces of hardware and software to be put together. But it also requires many other pieces of work to be done.
- [...]our SW&C projects are defined too narrowly and [are] defined in terms of components like Tier 1 centers, Tier 2 centers, Core software, etc. that are merely incidentals to the work that must be done to achieve the working environment for the future [...]
- [...] the components that are being built are functional ones. Each component starts off with lesser functionality (or non-existent at first) and each grows in capability and importance to meet the needs at each stage of the experiment and the schedule.
- [...] the end deliverables are indeed the entire environment for getting work done, living, discovering, educating and interacting
- Opportunities for sharing – between US Atlas and US CMS and between LCG and Europe and the US will hopefully become easier to identify and agree upon if we can lay out the work needed in a way that is more closely aligned to what activities and goals we need to accomplish.
- It is the LHC village with roads leading to many branches of globally oriented academic science.

DOE and NSF funding and "missing pieces"

US LHC Research Program: DOE + NSF funding for S&C projects

- Tier-1 and Tier-2 centers
  - equipment and operations
- some global services
  - user support, T2 support, etc
- CMS Core software engineering
  - US contribution to CMS effort



Berkeley workshop Nov 2002:

- "building the environment for a global collaborative science community"
- identify the "missing pieces"
  - 2 ITR proposal were scoped out: GECSR, DAWN
  - "Transition to production quality Grid"

Funding agencies: $50M project, DOE-NSF, international partnerships

Transition to Production-Quality Grid

LHC is Starting a Production Grid Service Around the World

DOE/NSF Partnership for Global Infostructure

Agreement on 5 principles:

I.Gaines, 4-Agency meeting at CERN March 21st, 2003

- The cost and complexity of 21st Century Science requires the creation of **advanced and coherent global Infostructure** (information infrastructure).
- The construction of a coherent Global Infostructure for Science requires **definition and drivers from Global Applications** (that will also communicate with each other)
- Further, **forefront Information Technology must be incorporated** into this Global Infostructure for the Applications to reach their full potential for changing the way science is done.
- **LHC** is a near term Global Application requiring advanced and un-invented Infostructure and is **ahead in planning** compared to many others.
- U.S. agencies must work together for effective U.S. participation on Global scale infostructure, and the successful execution of the LHC program in a **4 way agency partnership**, ***with international cooperation in view***.

21-Mar-2003 Partnerships for Global Infostructure

Physics + Computer Science/Information Technology Funding Agencies

"Work Packages" for LHC Computing

Facilities and Fabric Infrastructure

- U.S. Tier-1 and Tier-2 centers, U.S. University infrastructure

Distributed Computing Infrastructure

- Networks, throughput, servers, catalogs

Grid Services

- Middleware, "Virtual Organizations" support, end-to-end and higher level services, trouble shooting and fault tolerance, distributed science environment

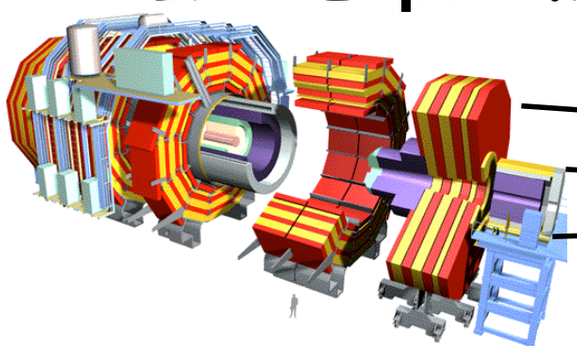
Experiment Specific Software

- Core software, frameworks, architectures, applications physics and detector support

Collaboratory Tools and Support

- Communication, conferencing, sharing, Virtual Control Room

Facility-Centric Components View



LHC Experiment

Online System

100-200 MBytes/s

Tier 0

CERN Computer Center > 20 TIPS

2.5 - 10 Gbits/s

Tier 1

Japan

UK

France

USA

...

Peta Scales!!

2.5 Gbits/s

Tier 2

Tier2 Center

2 Center

2 Center

2 Center

~0.6 Gbits/s

Tier 3

Institute

stitute

stitute

stitute

1 Gbits/s

Physics cache

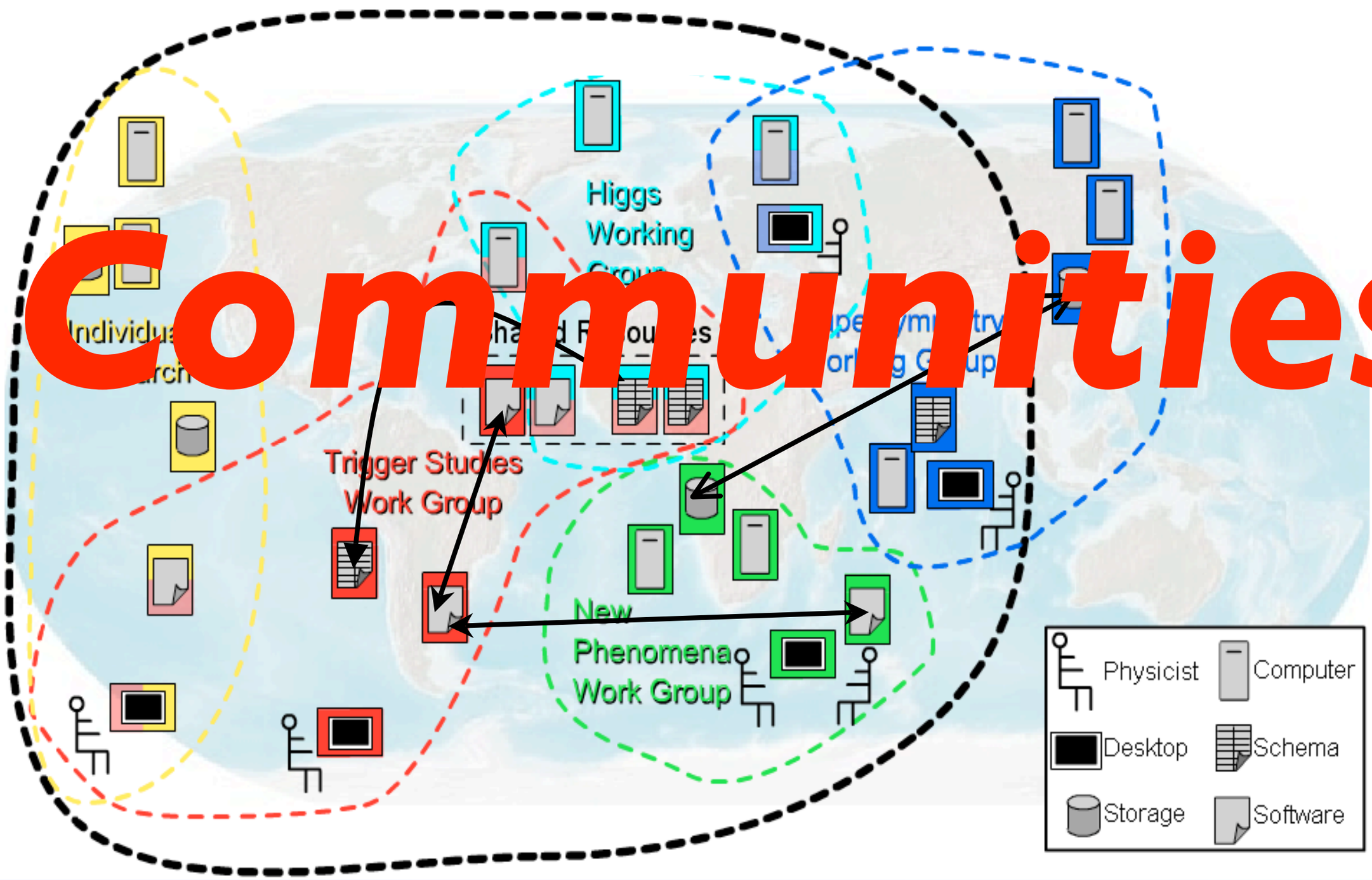
Tier 4

PCs, other portals

... Towards The Environment View: DAWN

Communities of Scientists Working Locally within a Global Context

Infrastructure for sharing, consistency of physics and calibration data, software



Communities!!

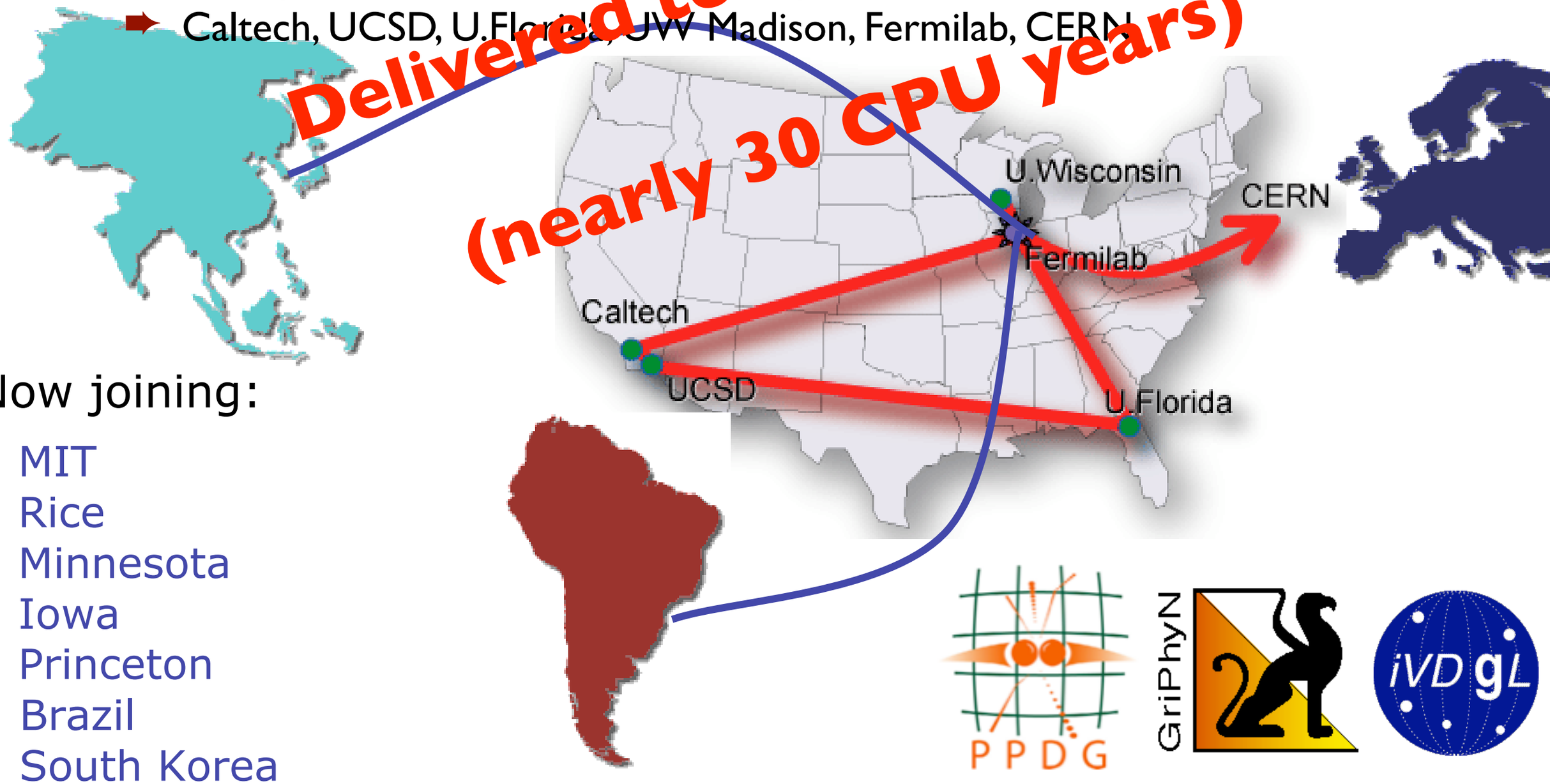
Grid Testbeds becoming Production Grids

Grid Testbeds for Research, Development and Dissemination!

- USCMS Testbeds real-life large Grid installations, becoming production quality
- Strong Partnership between Labs, Universities, with Grid (iVDGL, GriPhyN, PPDG) and Middleware Projects (Condor, Globus)
- Strong dissemination component, together with Grid Projects
- Caltech, UCSD, U.Florida, UW Madison, Fermilab, CERN

1.5 Million Events Delivered to CMS Physicists!

(nearly 30 CPU years)



- Now joining:
  - MIT
  - Rice
  - Minnesota
  - Iowa
  - Princeton
  - Brazil
  - South Korea

Where do we stand?

US CMS has a first instance of a persistent Grid infrastructure in place.

- This US CMS Grid is functional and is becoming a stable production environment for CMS scientists.

LHC Grids are based on the VDT, and US CMS has close ties to VDT team

- the distribution of Globus, Condor and other middleware components maintained by the VDT team
- will also underly the LCG software, and will likely be the basis for the proposed EGEE Grid infrastructure in Europe

Fermilab and US CMS develop, integrate and "showcase" Grid technology

- several examples for major "Grid related" technology advances
  - stable Grid running on US CMS IGT over months
  - TB/day data transfers between US CMS Tier-1 and Tier-2 and Tier-0
  - use of dCache at US CMS T2 for 10^34 pile-up running
  - and there are more examples

NSF M.Goldberg: Functional Grid Demos

Definition

- A series of functioning grids for use by LHC people and others- version zero in November designed to be used in X countries and handle Y data. Each succeeding version (~6 months) will multiply these numbers by N>>1.

Opportunity

- We are missing an opportunity to interest the world, not just us (example: possible GSF Agenda; Japan interest....) but this must be promoted correctly. We are 'doing' much of this but few know it.

Origin

- Resulting from 4-way NSF/DOE meetings and with CERN/EU and our joint "principles."

Features

- Illustrates Agency/LHC leadership in global grid development, told in ways designed to reach a large and important international audience.
- Adds value to particle physics.
- Tests computer science concepts and their robustness in a real global application.
- Aligns project contributors and their products in a common cause.
- Allows broader audience (science/geology/biology) to be contibutors/testers.
- Serves as important milestones in getting the LHC "done."
- Provides real world tests of functionality.
- Points to what is needed next.
- Tests contributors schedules
- Thus is a very important management tool
- Scientists working in this area will want to plug their stuff into it.
- Possible proposal for DOE/NSF.

DOE G.Crawford - Where is the $35M proposal?

LHC Consortium talks at DOE in April: Generated R.Staffin's interest

- request to quantify LHC computing challenge and Grid achievements
- indicated possible interest at the level of R.Orbach
- "is it true that LHC is at the fore-front? If you can proof and quantify, this is helpful for the field of HEP"

Last week phone conference with DOE/NSF

- follow up on 4-agency talks, next steps
- discussion on Grid demonstrations (Marv): CISE interest
- discussion on proposal to DOE: get in line for 2005
- I outlined a general thrust for such a proposal
  (after talking with a couple of people)
  and was tasked to draft a 3-pager outline this week

Open Science Grid

We propose to develop a plan

- and a proposal to the DOE and NSF

and to forge an organization and collaboration

- building upon the previously proposed "Open Science Consortium"

to build an "Open Science Grid" in the US

- for the LHC and other science communities

Goals and Scope:

- Develop and deploy services and capabilities for a Grid infrastructure that would make LHC computing resources, and possibly other computing resources for HEP and and other sciences (Run 2 etc) available to the LHC Science community,
  - as a functional, managed, supported and persistent US national resource.
- Provide a persistent Grid that peers and interoperates, interfaces to and integrates with, other national and international Grid infrastructrures
  - in particular the EGEE in Europe (which will provide much of the LHC Grid resources in Europe to the LCG)

The EGEE Mission

1. To deliver production level Grid services, the essential elements of which are manageability, robustness, resilience to failure and a consistent security model, as well as the scalability needed to rapidly absorb new resources as these become available, while ensuring the long-term viability of the infrastructure.

2. To carry out a professional Grid middleware re-engineering and development activity in support of the production services. This will support and continuously upgrade a suite of software tools capable of providing production level Grid services to a base of users which is anticipated to rapidly grow and diversify.

3. To ensure outreach and training effort which can proactively market Grid services to new research communities in academia and industry, capture new e-Science requirements for the middleware and service activities, and provide the necessary education to enable new users to benefit from the Grid infrastructure.

The NSF Cyber Infrastructure View

National Science Foundation

Blue Ribbon Panel on Cyberinfrastructure

Coordination (synergy) Matrix

Applications of information technology to science and engineering research

Cyber*infrastructure* in support of applications

Core technologies incorporated into cyberinfrastructure

Research in technologies, systems, and applications	**Development** or acquisition	**Operations** in support of end users

Infrastructure Layers

Between the Applications and the Computing Resources

Application Communities

Bioinformatics

Astrophysics SDSS

Run 2 CDF, D0

LHC Atlas, CMS, Alice

Applications of information technology to science and engineering research

Cyber*infrastructure* in support of applications

Core technologies incorporated into cyberinfrastructure

Resource Providers

General Facility for any Community e.g. TeraGrid

Facility Serving Multiple Communties e.g. Fermilab

Community Facility e.g. US CMS Tier-1 and Tier-2

University Facility e.g. UW Madison CS Condor

LHC has "bought into" this Grid Layer

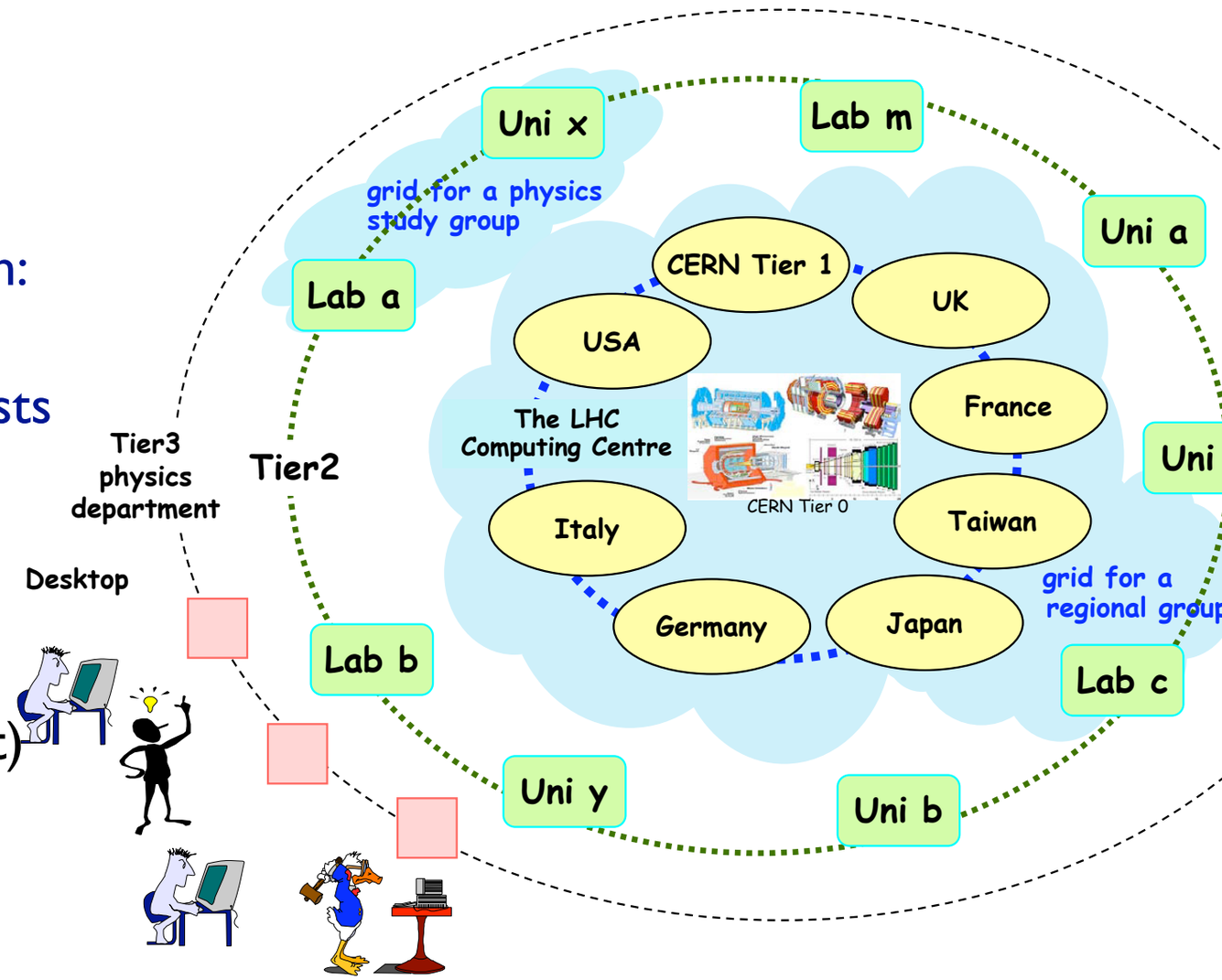
Provides "Virtualization": hiding useful functions behind an interface that conceals the details how they are implemented

- virtualization of computing and data services
- Grid of networked processing elements and data storage elements
- "middleware" provides the glue

Creates the environment the Applications live in:
Virtual Computing Service for Experiments
Ubiquitous Responsive Environment for Physicists

- distributed yet coherent computing
- coordinated and efficient sharing of geographically distributed resources
- conditional sharing issues of trust, policy, negotiation, payment)
- optimization of the resources
- invisibility of the local architecture
- partnerships and collaboration

Vision:
A Richly Structured, Global Dynamic System



Architecture — Grid -Application Layers

HEP Grid Architecture: (H. Newman)

Layers Above the Collective Layer

Physicist's Application Codes

- Reconstruction, Calibration, Analysis

Experiments' Software Framework Layer

- Modular and Grid-aware:
  Architecture able to interact effectively with the lower layers (above)

Grid Applications Layer
(Parameters and algorithms that govern system operations)

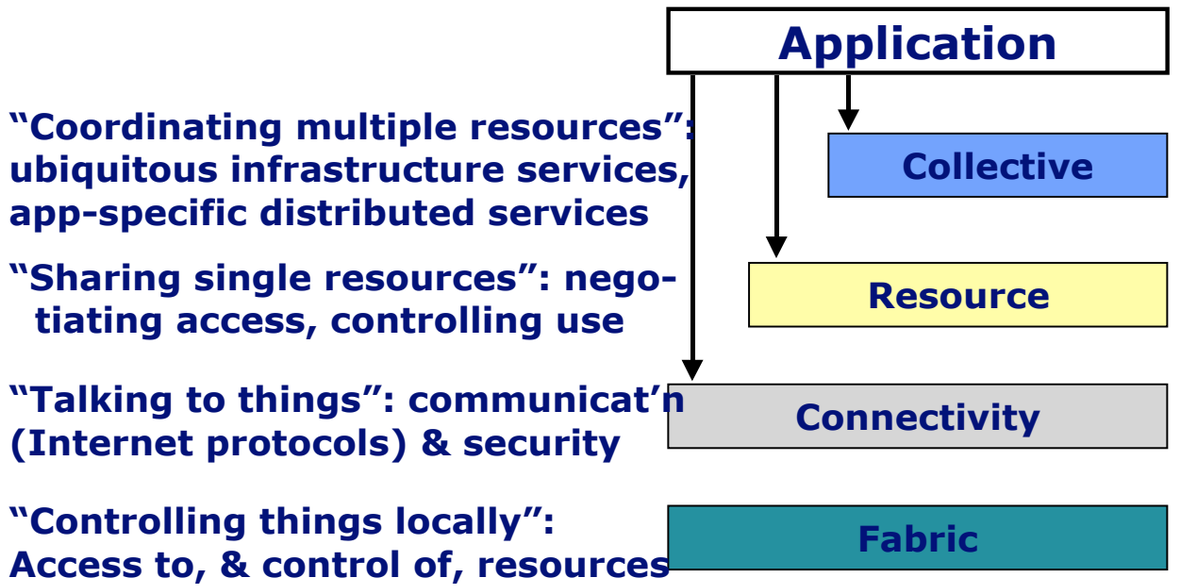
- Policy and priority metrics
- Workflow evaluation metrics
- Task-Site Coupling proximity metrics

Global End-to-End System Services Layer

- Workflow monitoring and evaluation mechanisms
- Error recovery and long-term redirection mechanisms
- System self-monitoring, steering, evaluation and optimization mechanisms
- Monitoring and Tracking Component performance

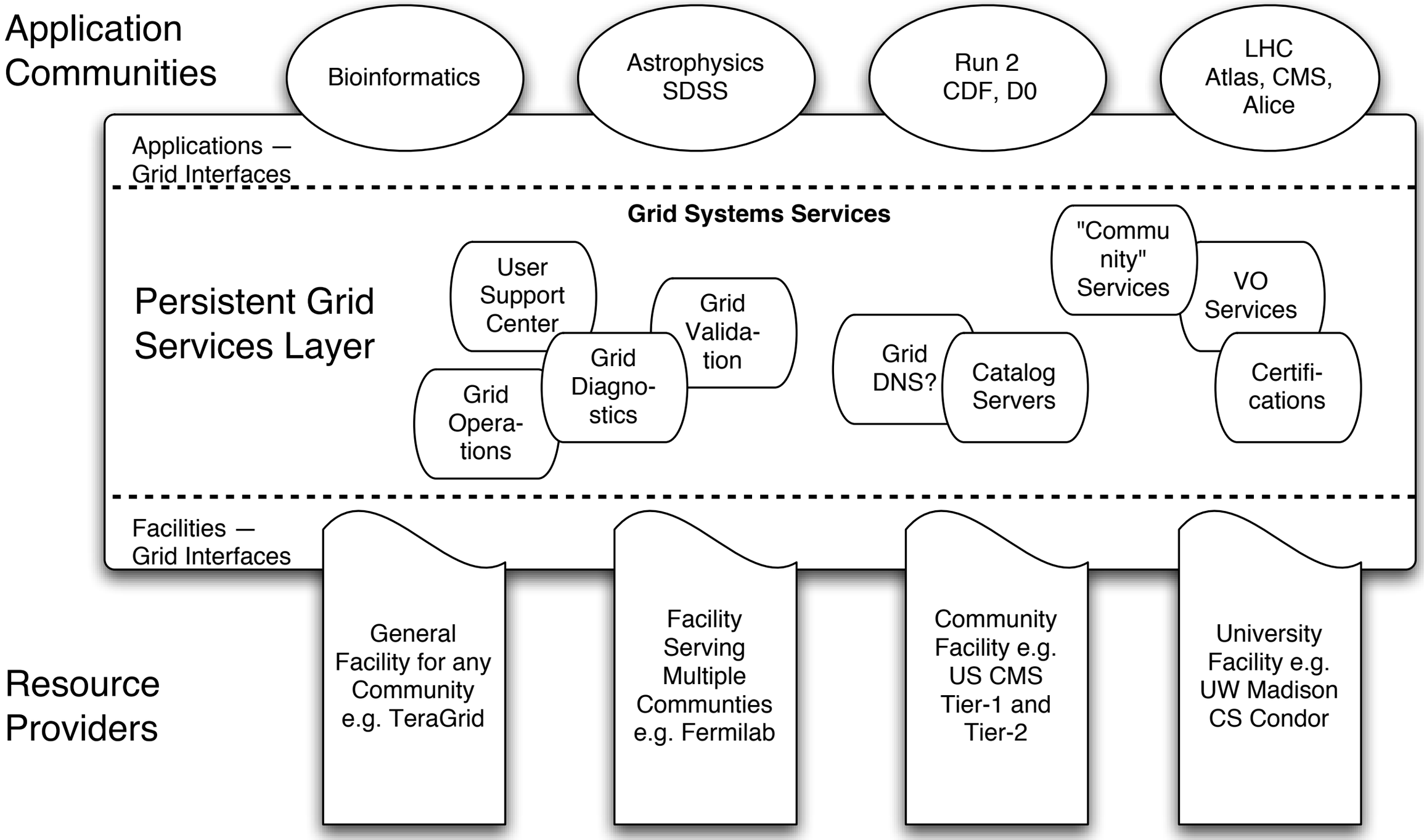
Layered Grid Architecture

(I.Foster et al.)



Infrastructure of Grid Layers

Grid Layer "Abstraction" of Facilities — Rich with Services!



This Project Should bring us There!

Scope out the layers between the Applications and the Facilities

- LHC already has identified funding for the fabric and it's operation

Work packages to acquire and/or develop enabling technologies as needed

- goal to enable "persistent organizations" like the national labs to provide those infrastructures to the application communities (CMS, Atlas, etc)
  - develop the "enabling technologies" that allow the fabric providers to function in a Grid environment, and the applications and users to seamlessly use it for their science
  - develop well defined interfaces
  - work on the technologies enabling end-to-end managed resilient and fault tolerant systems
- devise dependable "contracts"

Put up the initial operation infrastructure

Some of the Components

Persistent Grid Deployment and Operations

- Grid Operation and Support
- Policy and Management
- Grid Services
- Security
- Troubleshooting
- Quality Assurance and Regression testing

Middleware Maintenance and Support

- Middleware Hardening
- Interface Standardization

Community Application and System Integration and Support

- Application System Integration and Support
- International Grid Interoperability, Interfacing and Integration
- Trust and Accessibility
- User Support and Documentation

Dissemination and Support for non-Partners, Education and Outreach

A US Grid for the LHC

US CMS and US ATLAS leaders in facilitating this

many players -

- IT/CD divisions of the national labs, peering with the NSF
- Trillium, VDT, NMI, CS groups

pairing with "LCG + EDG + EGEE @ CERN + EU RCs" at the level of "US CMS + US ATLAS + VDT + Trillium @ FNAL + BNL + LBNL + "

building on that over the short, medium and long terms for the persistent infrastructure.

An infrastructure that the LHC projects - as representing the largest challenged community with proven functioning (small) grids to their name - can rely on, have confidence that ongoing operation and support models are in place for, and then get on with doing their science

And that we can offer others to join into, and be part of building up

Open Science Consortium

Open Science Consortium

Answer the call from U.S. funding agencies to address the breadth of LHC related computing needs in a overarching collaboration including more than U.S. Atlas and U.S. CMS

Propose an "Open Science Consortium" to provide a framework under the umbrella of the U.S. LHC Research Program, as a collaboration between the existing and currently proposed U.S. LHC related projects, LCG and other international projects

- to enable and address collaboration to more than across the US LHC experiments: Alice, LHCb, RunII, BaBar?, BTeV? , ...
- address remaining missing capabilities for U.S. LHC computing through joint initiatives, by developing common future proposals for enhancing "cyberinfrastructure" for our community as driver for other sciences
- to disseminate the LHC technologies to other sciences

er

onsortium

ment and Dissemination of

Data Handling and Analysis

ndertaken as an outreach of the U.S.
p by the US Tier 1 laboratories ??)

sting and currently proposed U.S. LHC
ıysics experiments, other domain

e the Open Science Consortium as an
ınologies being developed as part of its
ı and grid projects.

The Open Science Consortium will through joint initiatives share, develop, or otherwise acquire, necessary missing advanced information technology capabilities to advance the adoption and deployment of Global Application Grids for the scientific community.

Open Science Consortium

Open Science Consortium

Answer the call from U.S. funding
related computing needs in a ove
U.S. Atlas and U.S. CMS

Propose an "Open Science Cons
umbrella of the U.S. LHC Researc
existing and currently proposed l
international projects

- to enable and address collab
experiments: Alice, LHCb, Ru
- address remaining missing ca
joint initiatives, by developing
"cyberinfrastructure" for ou
- to disseminate the LHC tech

Charter

Open Science Consortium

A Collaboration for the Advancement and Dissemination of Global Infostructures for Scientific Data Handling and Analysis

The Open Science Consortium is an initiative undertaken as an outreach of the U.S. LHC Research Program, (?? with co-sponsorship by the US Tier 1 laboratories ??)

with the mission of collaboration across the existing and currently proposed U.S. LHC related projects, other US particle and nuclear physics experiments, other domain scientists and international peer projects.

The US LHC Research Program will actively use the Open Science Consortium as an avenue for dissemination of the information technologies being developed as part of its software and computing and associated research and grid projects.

The Open Science Consortium will through joint initiatives share, develop, or otherwise acquire, necessary missing advanced information technology capabilities to advance the adoption and deployment of Global Application Grids for the scientific community.

Partners, Stakeholders, Collaborators?

Partners:
1. DOE HENP Laboratories serving multiple user communities:
a. Fermi National Accelerator Laboratory
b. DOE Science Grid (LBNL)
c. BNL
d. TJNAF – if LQCD

2. Middleware CS Partners
a. University of Chicago
b. University of Wisconsin Madison
c. University of Tennessee (Markus Lorch)
d. LBNL – DOE Science Grid
e. Stonybrook/BNL consortium

3. Stakeholders:
a. Run II experiments – CDF, D0,
b. US ATLAS, US CMS, US ALICE Software and Computing projects
c. SDSS, NVO
d. LQCD – I think we should try and have them as stakeholder partners
e. BTeV
f. SciDAC Advanced Accelerator Simulation
g. Need other science …

4. Collaborators (Unfunded)
a. Particle Physics Data Grid
b. IVDGL

Action Items

-
-
-
-
-
-
-
-
-
-

We got to be ready on day 1!

Commissioning a Early Physics with

Kerstin Hoepfner

RWTH Aachen, Institute of Physics IIIA

On Behalf of the CMS Collaboration

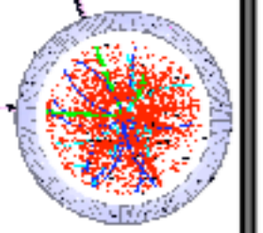
The Start-up LHC Scenario

PHASE 1: LHC commissioning (T_0 to T_0 + 4 months)

- LHC: Set-up machine. Start with one beam. Colliding beams and slowly increase # bunches and **L**. Collisions at L > 5 x 10^{32} at 25, 75 ns bunch spacing.
- CMS: Muon halo triggers, catalog detector problems, synchronization, debug data handling, record first collisions

→ This talk part I

+ 2 pileup events i.e. 10^{33} and 25ns bunch spacing

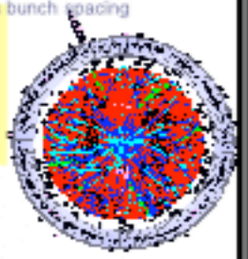


pixel **Phase2: Shutdown**

PHASE 3: First physics run (T_0 + 7 mo.→ T_0 + 14 mo.)

- LHC: 25 ns and **L** ~ 10^{33} cm^{-2} s^{-1}
- CMS: Physics run, max. efficiency aiming for 5-10 fb^{-1}
- ~2 events per BX → This talk part II

+ 20 pileup events i.e. 10^{34} and 25ns bunch spacing

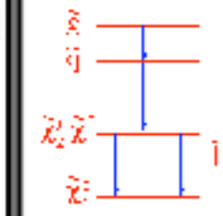


PHASE 4+n: High luminosity running

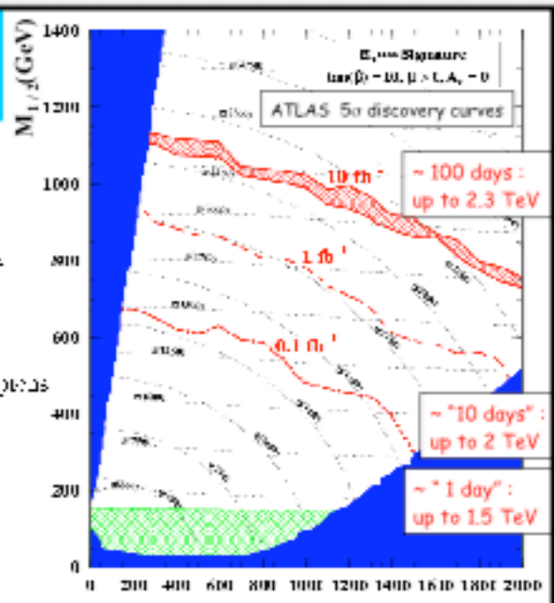
→ See talk by J.Rohlf on high luminosity

SUSY

Large cross-section for squark and gluino production



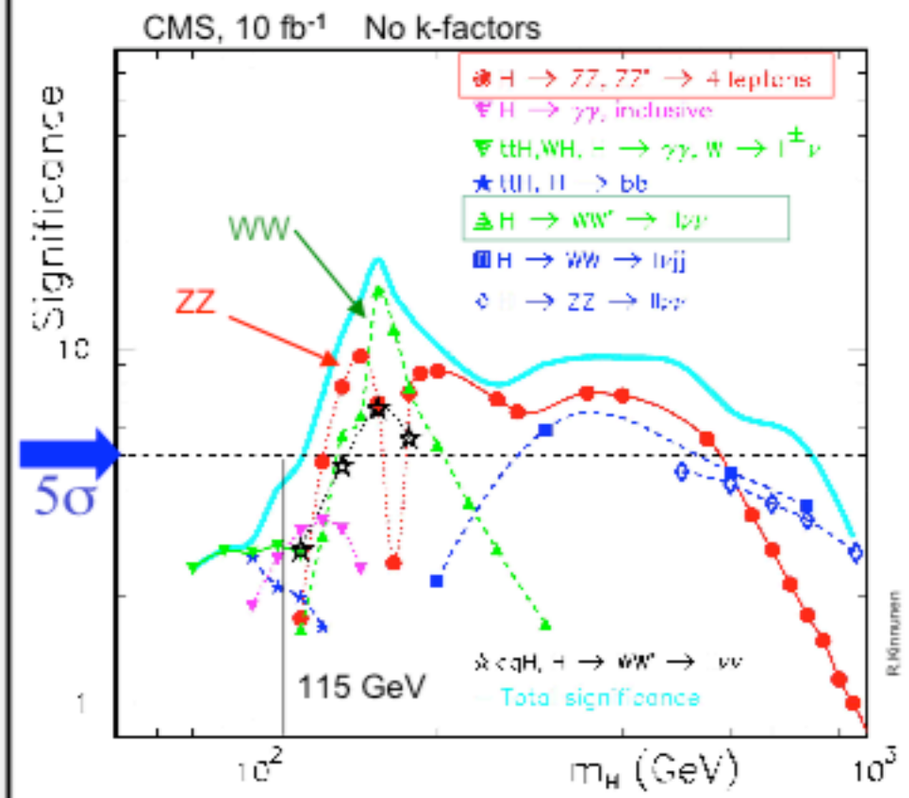
Decay chain leads to
- **high p_T jets**
- **large missing E_T**
- **isolated leptons**



Discovery of SUSY is easy for masses below 2 TeV

May 2nd, 2003 Ivor Fleck ATLAS, Early physics reach

SM Higgs Discovery Potential with 10 fb^{-1}



In the following:

H → 4 leptons,
m_H = 120….500 GeV

H → WW → lνlν
m_H = 110….200 GeV

K.Hoepfner, RWTH Aachen Early Physics with CMS, LHC Symposium May 2003 24

Conclusions

- **Commissioning of detector challenging**
- **procedures are being developed now**
- **within first days:**
  - **Alignment of central detector using muon tracks to < 2 mm**
  - **Calibration of EM using Z → ee to 0.6 %**
- **Impact of staging: Need ~ 10 - 15 % more integrated luminosity**

Physics results within first year:
- **Higgs boson may be discovered over full mass range (low mass region very challenging)**
- **MSSM Higgs likely to be seen**
- **LHC is factory for SUSY particles, discovery immediately first year reach is squark masses up to 2 TeV**

Looking forward to many interesting physics analyses

May 2nd, 2003 Ivor Fleck ATLAS, Early physics reach